

Title	内面状態の推定と対話戦略適応に基づく知的対話処理に関する研究
Author(s)	長澤, 史記
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19928
Rights	
Description	Supervisor: 岡田 将吾, 先端科学技術研究科, 博士

氏 名	長澤 史記
学 位 の 種 類	博士（情報科学）
学 位 記 番 号	博情第 550 号
学 位 授 与 年 月 日	令和 7 年 3 月 21 日
論 文 題 目	Research on intelligent dialogue processing based on inner state estimation and adaptive dialogue strategies
論 文 審 査 委 員	岡田 将吾 北陸先端科学技術大学院大学 教授
	池田 心 同 教授
	白井 清昭 同 准教授
	宮田 一乗 同 教授
	中野 有紀子 成蹊大学 教授

論文の内容の要旨

The ultimate goal of this research is to realize a computer system that can connect users to appropriate functions and services by empathizing with their feelings and interests through adaptive interactions and eliciting their inner needs. As a method for eliciting information from people, listening dialogue systems that use dialogues such as interviews and interviews have been actively studied. In listening dialogue systems, it is important to realize empathic dialogue, in which the system recognizes the inner state of the user and motivates the user to participate in the dialogue.

To realize such empathic dialogue processing, it is necessary to realize a mechanism in which the system proposes or changes topics according to the user's willingness, which is the user's inner state of "wanting to talk".

First, a machine learning model was developed to estimate the interviewee's willingness (desire to disclose information they have) based on their posture during the dialogue and the prosodic features of their speech utterances. Machine learning models using Random Forest and LinerSVM were trained. A method was developed to reduce the impact of individual differences in multivariate features on estimation accuracy for first-time interviewees who do not have the information necessary for normalizing multivariate features. Using the interview corpus collected through the dialogue experiment, this study evaluated the accuracy of the estimation of willingness by cross-validation, and found that the method correctly estimated high and low willingness with an accuracy of up to 72.8%.

Next, we introduced an adaptive dialogue strategy using this willingness recognition model and implemented it in a dialogue robot. In the adaptive dialogue strategy, if the estimated willingness is high, the robot continues the topic of the previous question, and if the estimated willingness is low, the robot switches the topic. Question selection by topic continuation/change was realized by exploring a pre-constructed question graph (a tree graph in which questions are arranged based on topic relevance). A dialogue experiment was conducted with 27 participants to evaluate the effect of adaptive dialogue strategy. The dialogue experiments were compared between the proposed

adaptive dialogue strategy system and a random strategy system with random topic continuation/transition. The experimental results confirmed that the adaptive dialogue strategy gave users the impression that they were listening with more interest, and also significantly increased the number of utterances with high willingness. This showed that even with less-than-perfect estimation accuracy, it is possible to motivate users to speak through adaptive dialogue strategy.

In order to improve the accuracy of multimodal inner state estimation and to analyze the accuracy of attitude estimation due to individual differences in multimodal features and the sensing environment, this study worked on refining the attitude estimation using the pre-built external corpus Hazumi1911.

We trained and evaluated a model that added biodata and facial landmark features in addition to prosody and posture features. As a result, the addition of the features used improved the accuracy, and the individual differences in estimation accuracy decreased. We evaluated the accuracy of models trained on two different corpora with different sensing environments. The models were trained on the Hazumi1911 corpus and evaluated on the accuracy on data outside the corpus (a newly collected corpus of interview dialogues). The results showed that the accuracy of the models on data outside the corpus decreased, and the difference in accuracy between individuals also increased. However, the decrease in accuracy was smaller for models that used more features, and the difference in accuracy between individuals was also smaller.

A question generation method based on a large-scale language model (LLM) was proposed for the purpose of making adaptive dialogue strategy applicable to arbitrary topics. We implemented an improved interview robot system that incorporates updated willingness recognition model with extended features used and LLM-based adaptive question generation. The results of a dialogue experiment with 30 interviewees showed that the degree of self-disclosure of the interviewees improved when the adaptive strategy was used compared to the random strategy.

In summary, this thesis presents the results of an analysis of multimodal inner state estimation based on nonverbal information during dialogue, the implementation of an adaptive dialogue system, and its impact on dialogue. Our results show that an adaptive dialogue strategy increases user willingness, promotes self-disclosure, and lead to better interviews even with user adaptation using imperfect inner state estimation models.

These results will lead to new applications of dialogue technology through interview techniques that promote self-disclosure of the subject and elicit deeper narratives. For example, by eliciting the user' s unspoken feelings and inner narratives, it will be possible to improve counseling and service recommendations. The realization of such assistant technology that empathizes with the user' s inner world and proactively suggests solutions to the user' s problems will greatly improve future human-computer interaction.

Keywords: Sentiment Analysis; Physiological Signal Processing; Machine Learning; Multimodal Signal Processing; Dialogue System.

論文審査の結果の要旨

本論文は、マルチモーダル内面状態推定に基づくヒューマン・ロボット・インタラクション (HRI) において、マルチモーダル対話的インタラクションにおけるユーザ適応の効果を体系的に明らかにすることを目的とした。対話において、ユーザの内的な状態を認識し、会話の継続を促す共感的対話処理を実現し、ユーザ適応の効果を検証した。

まず、対話者の発話意欲（持っている情報を開示したいという欲求）をオンラインで推定するモデルを構築・評価し、72.8%の精度で発話意欲レベルの高低を推定できることを示した。また、多変量特徴量の正規化に必要な情報を持たない初対面者に対しても、個人差の影響を低減する手法を導入した。さらに、この意欲推定モデルと、意欲レベルに応じて質問を掘り下げたり話題を転換したりする適応的質問戦略を実装し、ヒューマノイドロボットに搭載した。

提案した適応的質問戦略を備えたシステムと、ランダムに話題を継続・転換するランダム戦略を備えたシステムを比較した結果、適応的対話戦略を用いることで、ユーザに「興味を持って聞いている」という印象を与え、高い意欲を持って発話する回数を有意に増加させることができた。これは、推定精度が完全でなくても、適応型対話戦略によってユーザの発話を促進できることを示している。次に、マルチモーダルな内的状態推定の精度向上を目的として、韻律や姿勢の特徴量に加え、生体情報や顔のランドマーク特徴を統合した推定モデルを学習・評価した。これにより、推定精度が向上し、個人差の影響が減少することを確認した。さらに、適応型対話戦略を任意の話題に適用するため、大規模言語モデル (LLM) に基づく質問生成手法を提案した。改良した内面推定モデルと LLM を活用した対話システムをインタビューロボットに実装し、30 名のインタビュー対象者との対話実験を実施した。その結果、改良された適応的質問戦略を有するシステムを使用することで、インタビュー対象者の自己開示の程度を有意に向上させることができた。

本論文は、マルチモーダル内面状態推定モデルの構築にとどまらず、それを活用したオンライン対話システム（ロボット）を実装し、対話における内面状態推定の効果を実証した点に学術的な意義がある。また、本論文の一部は、Affective Computing 分野のトップランクジャーナル IEEE Transactions on Affective Computing に採録されており、今後のヒューマン・コンピュータ・インタラクション分野に貢献する成果を有している。以上より、本論文は学術的貢献が大きく、博士（情報科学）の学位論文として十分な価値があると認める。